Modeling nonresponse in multiwave panel studies using discrete-time Markov models

Taris, T.W.

published in
Quality and Quantity
1996

DOI (link to publisher)
10.1007/BF00153987

document version
Publisher's PDF, also known as Version of record

Quality & Quantity **30**: 189–203, 1996.

Modeling nonresponse in multiwave panel studies using discrete-time Markov models

TOON W. TARIS
Kurt Lewin Institute/Free University Amsterdam, Department of Social Psychology, van der Boechorst-Straat 1, 1081 BT Amsterdam, The Netherlands

Abstract. Nonresponse is of major concern to social scientists, due to the possibility of selectivity: not all groups in the population are equally represented in the final sample, when some groups have a larger probability to be in the sample than others. It is dangerous to base conclusions on such biased samples. Therefore, it is of importance to study nonresponse patterns. First it is shown that a decreasing nonresponse for every successive wave indicates that nonresponse is selective to a degree. Successively we discuss how Markov models can be used to get some idea of the seriousness of this bias in the sample, by examining how many chains are needed to reproduce the observed pattern of nonresponse acceptably well, and what the probability is that members of these chains will participate in a particular wave of the study. A small application is given, after which the implications of the findings are discussed.

1. Introduction

Nonresponse – not being able to obtain data from all subjects who initially were meant to be in the sample, either due to explicit refusal to participate or to other reasons – is an issue that is of major concern to us all. It is often the case that nonresponse rates are as high as 30 to 40 per cent. For example, Goyder (1987), in his extensive review of nonresponse rates in mail, face-to-face, and telephone surveys, reports that the nonresponse in these studies on average amounted to 41.6, 32.7, and 39.8 per cent, respectively. Similar findings, though based on a much smaller number of studies ($N = 45$), were reported by Hox and De Leeuw (1993).

A high response rate is considered to be important for three reasons. First, it is obvious that if the number of subjects that does not want to cooperate in a study is large, data collection becomes a *costly matter*. Many subjects need to be approached before one's sample is large enough to suit one's needs. This applies even more to longitudinal research, where the same set of subjects is approached at least twice. The subjects who participate in the first wave, but who drop out in a later stage, increase the costs of the study to no avail.

Secondly, if nonresponse is high, there are few subjects left to test one's hypotheses. This means that nonresponse leads to a *low statistical power*, due to small sample size. If one expects a large nonresponse, this disadvantage can be countered by approaching many subjects – but this, again, increases the costs of data collection.

In the remainder of this paper we will focus on the third reason why a high response rate is considered to be important. It has often been noted that a high nonresponse bears the risk that nonresponse is *not random* (or *selective*): groups in the population differ concerning the chance that they will be in the sample, leading to differences between the target population and the sample to be analysed. For example, it has been shown that inhabitants of large cities, women, youngsters, elderly, low SES, and low education subjects tend to be underrepresented in samples (cf. the review in Van de Pol, 1989). Of course, conclusions that are based on analyses conducted on such a biased sample cannot readily be generalised to the population. Restriction of range-effects can be expected, and if variables of interest are related to variables that 'cause' nonresponse or dropout, the estimates of their effects on other variables will be biased. Hence, nonresponse is a major threat to the validity of studies.

This applies even more to longitudinal research, as here nonresponse tends to accumulate. Subjects who refuse to participate in the first wave of a study are usually not approached for the second wave, and so on. This can greatly reduce the size of the sample. If R denotes the (constant) response probability that a subject will participate in any particular wave of the study and w is the number of waves, R^w represents the fraction of subjects who are still in the study after w waves. For example, if $R = 0.8$ and $w = 4$, only 41 per cent of the subjects will be in the study after four waves. The review in Van de Pol (1989) shows that this is not an exceptionally high attrition rate.

Of course, the nonresponse is not necessarily the same for all waves of a particular panel study. A frequently encountered observed fact is that nonresponse decreases with every wave of the panel. For example, Taris *et al.* (1993) report for a three-wave panel study a nonresponse of 37, 20, and 11 per cent for each successive wave, compared with the previous wave. Such a decreasing nonresponse is *not* the good omen it is often considered to be. On the contrary, as we will show below, such a decreasing nonresponse is a strong indication that the sample is getting less representative for the target population, due to selective nonresponse. Below this phenomenon is discussed more fully. Successively we will show how patterns of nonresponse can be studied by means of Markov modeling.

2. Example: a decreasing non-response rate spells trouble

We will first provide a small synthetic example that shows how selective nonresponse operates, and what the implications can be. Let a population consist of two equal-sized groups (A and B), that differ concerning their response probabilities R_a and R_b. For example, let group A consist of – say – female low-SES inhabitants of large cities, while group B consists of well-educated male high-SES inhabitants of suburbs. Let their response probabilities be 0.6 and 0.9, respectively. Finally, assume that members of both groups have an equal chance to be asked to participate in the study. How many members of both groups need to be approached to obtain a first-wave sample of 1000 subjects, and how will the sample be composed after a number of waves, if R_a and R_b are constant over time?

The first question can be answered by solving the equations

$$R_a\mathrm{A} + R_b\mathrm{B} = 1000, \tag{1a}$$

$$\mathrm{A} = \mathrm{B} \tag{1b}$$

which shows that a total number of 1334 persons (667 of each group) must be approached before we have a 1000-subject sample. Hence, the total nonresponse for the first wave is slightly more than 25 per cent (25.04%). Of the 667 members of group B, 600 subjects agreed to participate; however, only 400 members of group A are in the sample. The A-to-B ratio A/B is, therefore, 0.67: for each member of A, there is 1.5 member of group B in the sample. Table 1a shows how these figures change over time for a four-wave panel study.

Clearly, group B becomes more dominant with each wave. At the fourth wave there are 5 members of group B for every member of group A. Obviously, after four waves the sample is anything but representative for the target population.

Table 1b shows that the rate at which B becomes dominant in the sample depends on the difference between R_a and R_b. If this difference is 0.1, selective nonresponse creates a situation in which the A-to-B ratio is 0.59 after four waves. For every member of group A, there is 1.7 member of group B present. Though this bias is not exactly small, it is much lower than was the case with $R_a = 0.6$ and $R_b = 0.9$. Clearly, the difference in response rates is very important with respect to the degree to which selective nonresponse is a problem.[1]

A *decreasing response.* The interesting fact is that in the example the nonresponse of every wave, compared to its predecessor, decreases. For

Table 1. Different response probabilities affect the composition of the sample, number of waves = 4

Table 1a: $R_a = 0.6$, $R_b = 0.9$

	Wave 1	Wave 2	Wave 3	Wave 4
Group A	400	240	144	87
Group B	600	540	486	438
Total sample	1000	780	630	525
Ratio A/B	0.67	0.44	0.30	0.20

Table 1b: $R_a = 0.7$, $R_b = 0.8$

	Wave 1	Wave 2	Wave 3	Wave 4
Group A	467	327	229	160
Group B	533	426	341	273
Total sample	1000	753	571	433
Ratio A/B	0.88	0.77	0.67	0.59

$R_a = 0.6$ and $R_b = 0.9$, the percentages nonresponse are 25.0, 22.0, 19.2, and 16.7, respectively; for $R_a = 0.7$ and $R_b = 0.8$ we find nonresponse percentages of 25.0, 24.7, 24.3 and 24.0. This is because group B – that constitutes a larger part of the sample with every wave – has a lower nonresponse rate than A. Indeed, as the number of waves increases, the nonresponse of the total sample comes closer to R_b with every wave.

Now compare the results in Table 1a with the situation in which the response probability R is the same for all subjects, for example, $R = 0.75$. To obtain a sample of 1000 persons at wave 1, we would still need to ask a total of 1334 persons to participate. At the second wave the sample size would be 750 subjects, at time 3 there would still be 563 persons, and, finally, at wave 4 there would be 423 subjects. The interesting fact is that the sample size in the R-homogeneous sample decreases *faster* than in the R-heterogeneous sample: at wave 4, there is a full 102 persons more present in the heterogeneous sample than in the sample with a constant response rate of 0.75. Again, this result is strongly dependent on the difference in response probabilities: if we take $R_a = 0.7$ and $R_a = 0.8$, the difference in sample size will only be 10 subjects at the fourth wave.

Hence, this example shows that in the heterogeneous group the nonresponse decreases; compared to the R-homogeneous group this leads to a larger sample (which is a desirable result, applying the rule of thumb that nonresponse should be as low as possible); the sample, however, is *less* representative for the target population than the heterogeneous sample. The interesting paradox here is, of course, that a relatively high total attrition goes together with a high representativity. Stated differently, a low nonresponse does *not necessarily* warrant representativity; it is the nonresponse *pattern* that informs us whether nonresponse is approximately random, or that there are good reasons to assume that we are dealing with a biased sample (Taris, 1994).

3. Studying nonresponse patterns as a Markov process

A more formal way of studying nonresponse patterns is to think of nonresponse and dropout as the result of the operation of a *Markov process* (Anderson, 1954; Bartholomew, 1981; Coleman, 1968; Markus, 1979). In its simplest form, a Markov-process is a probabilistic process occurring in discrete time, with regard to the score of a subject on a discrete (qualitative) variable X. The score on X should be available for all w waves of the panel study. In the case of nonresponse, X is a dichotomous variable with categories 0 = 'still in the study' and 1 = 'dropped out'. It is assumed that drop outs are not contacted anymore, i.e., once one is in state 'dropped out' one remains in that state.

A *first-order* Markov process is completely described by the vector $\boldsymbol{P}$ of probabilities that the subject belongs to state j (j = 1, . . . , p) and the $p \times p$ matrix of w to $w + 1$ transition probabilities $\boldsymbol{O}$ (Bartholomew, 1981; Chung, 1967; Freedman, 1970). Generally, for every wave w, $\boldsymbol{P}$ can be computed using

$$\boldsymbol{P}_w = \boldsymbol{P}_0 \times \boldsymbol{O}^w \tag{2}$$

assuming that the waves are equally spaced in time and with $\boldsymbol{P}_0$ the fraction of the sample that belongs to x_1 and x_2 at the first wave. Take, for example, group A in Table 1, with response probability $R_a = 0.6$. The first-order transition matrix $\boldsymbol{O}$ that corresponds with this probability is

$$\begin{bmatrix} 0.6 & 0.4 \\ 0 & 1 \end{bmatrix},$$

i.e., the probability to remain in state x_1 ('still in the study') at wave w, given that one was in the study at $w - 1$ is 0.6, the likelihood to experience a transition to x_2 at time $w + 1$ ('dropped out') given that one is in x_1 at w is 0.4, the probability to experience transition to x_1 given x_2 is zero, and the probability to remain in x_2 once one belongs to x_2 is 1 (i.e., x_2 is an *absorbing* *state*). This example shows that the operation of a constant response probability R can be described as a time-invariant first-order Markov process, i.e., the simple w to $w + 1$ transition matrix – which is constant over waves – and the vector $\boldsymbol{P}_0$ are sufficient to describe the attrition process at any given point in (discrete) time: at any w it is known how many subjects are in the study, and how many already have dropped out.

It is also possible that the observed nonresponse pattern does not fit the hypothesis that it is the result of the operation of a first-order Markov process. It could, for instance, be the case that a second- or higher-order Markov chain is responsible for the pattern of nonresponse (i.e., $\boldsymbol{P}_w$ cannot be predicted from $\boldsymbol{O}$ and $\boldsymbol{P}_{w-1}$ only; we need more information to do this, for example, also information about $\boldsymbol{P}_{w-2}$). However, for simplicity such higher-order Markov processes will not be considered here.[2]

It could also be the case that the observed pattern of nonresponse is the result of the simultaneous operation of two or more Markov chains (a *mixed* Markov model). Consider the synthetic example presented above once again. Obviously, the data in this example can be described in terms of the operation of two first-order time-invariant Markov chains, with different transition matrices $\boldsymbol{O}$. $\boldsymbol{O}_a$ was already presented above, and the reader will be able to infer $\boldsymbol{O}_b$ for him/herself.

The interesting question is, of course, whether the structure of the process that generated the nonresponse can be inferred from the nonresponse pattern. If this structure would be known, it would give us a handle to make inferences about the degree to which nonresponse is selective (i.e., how heterogeneous the sample is: as we have learnt from the example, the difference in response probabilities as well as their magnitudes makes a big difference), how biased the sample is after w waves, and what the size of the sample would be. Such information would be very handy in assessing the degree to which conclusions drawn using the sample can be generalised to the target population.

The mixed Markov model. The mixed Markov model assumes that each subject belongs to one of one or more subpopulations. Membership status h ($h = 1, \ldots, H$) is constant for all w occasions. Each member of a subpopulation h can belong to one of one or more 'chains', that is, a part of the (sub)population that has the same dynamics. A proportion $\tau_{s|h}$ in subpopulation h belongs to chain s. The proportion of the total sample that belongs to subpopulation h and chain s at wave 1 is $\gamma_h \tau_{s|h}$. Now, the number of subjects in cell P_{ijk} in the three-way table Wave 1 by Wave 2 by Wave 3 (i.e., the proportion of subjects in subgroup h that belongs to i at wave 1, j at wave 2, and k at wave 3) is given by

$$P_{ijk} = \sum_{h=1}^{H} \gamma_h \sum_{s=1}^{S} \tau^1_{i|sh} \tau^2_{j|sh} \tau^3_{k|sh}. \tag{3}$$

Maximum-likelihood estimates of γ and τ can be obtained using a version of the EM-algorithm (Van de Pol and Langeheine, 1989). The fit of the model can be tested by means of a chi-square distributed likelihood-ratiotest. As usual, this test compares the observed cell frequencies with the frequencies as expected on the basis of the estimated model parameters. It is therefore also possible to compare the fit of several models to the data, and select the model that gives the best fit, relative to the number of degrees of freedom.

4. Application

The theoretical discussion above will be supplemented with a small empirical example, using data from a three-wave study on the social integration of young adults. At the first wave (fall/winter 1987/88) a stratified sample of 1775 men and women was interviewed, equally divided by gender and birth cohort (1961, 1965 and 1969). At the first wave they were extensively interviewed about behaviour and attitudes concerning several life domains (employment, family formation, school). Two years later (1989/90) they received a mail questionnaire that updated the information with regard to the developments on the life domains (i.e., behaviour). Finally, a third wave – which was almost identical to the first: again, the subjects were interviewed by trained interviewers, and so on – was conducted during fall/winter 1991/1992.

The nonresponse percentages were 37, 20, and 11 for each wave, respectively (Taris *et al.*, 1993). Clearly, according to our expose above, the nonresponse could well be selective in this case, although the data collection was conducted extremely carefully, and great effort was spent in trying to reduce nonresponse. For example, subjects who cooperated in a particular wave received a Fl.10-gift coupon and a summary of the main results of the study. Great care was taken to trace subjects who had moved since the last interview, and so on (see Dijkstra, 1993).

We estimated several models that could account for the nonresponse pattern in this study. The first model of interest is the first-order Markov chain with constant transition probabilities. If this model fits the data acceptably well, it would imply that the nonresponse is constant over waves, i.e., this would comply with the situation in which all subjects have an equal and random chance of dropping out: there are no reasons to expect selective nonresponse on the basis of such a time-invariant pattern.

The second model is a variation of the first. Again, this is a simple first order Markov chain, but now we allow the transition probabilities to vary per wave. In fact, this is a model that does not restrict the data in any way: for every wave a separate transition matrix $\boldsymbol{O}$ is computed, and the model has no degrees of freedom. This model is included for expository purposes, not for its theoretical interest. In the remainder we will call this model the *chaos* model, because it assumes that nonresponse rates vary per wave, and that there is no particular mechanism behind this variation. I.e., nonresponse is not systematically determined by particular factors, but simply due to random variation.

The third model involves two Markov chains that operate at the same time (the mixed model). Two submodels will be considered. The first is a two-chain model in which the transition probabilities are constant over time for both chains, i.e., there are two first-order Markov chains in operation. The second submodel is a special case of the first, namely the *mover-stayer model* (Blumen *et al.*, 1966). According to this model, one of the transition matrices $\boldsymbol{O}$ is the *identity matrix*, i.e., a matrix with zeroes in the off-diagonal elements while the diagonal elements are equal to one. This in effect means that the members of such a stayer chain will never experience any transition at all; they stay where they are. The other transition matrix is not restricted, hence, members of this mover chain may experience transitions from one state to another.

These latter two models are of most interest, because they comply with the theoretical notions outlined in the synthetic example presented above. Given the current nonresponse pattem, we expect the two-chain models to have a considerably and significantly better fit to the data than the first-order Markov chain with constant transition probabilities.

The models were estimated by means of the PANMARK 2.2 program by Van de Pol *et al.* (1990). This program estimates the parameters of a wide range of Markov models, including mixed-Markov and state-of-the-art latent class models (Langeheine and Van de Pol, 1989; Poulsen, 1982; Van de Pol, 1989; Van de Pol and Langeheine, 1989; Wiggins, 1973). The data set is presented in the Appendix.

Results

Comparison of models. Table 2 presents the fit of the models. Clearly, the first-order Markov chain with constant transition probabilities cannot account

Table 2. Comparison of models

	Model description	LR	df	p
One chain	Time-invariant model (R the same for all waves)	1113.85	4	0.000
	Time-varying model (R not the same for all waves)	0.000	0	Undefined
Two chains	Time-invariant mover-stayer model (R the same for all waves)	0.177	2	0.91
	Time-invariant two-chain model (R the same for all waves)	0.169	0	Undefined

for the data. It is strongly rejected (likelihood-ratio LR with 4 degrees of freedom is 1113.85: $p < 0.001$). This result implies that the attrition is not constant over waves; the variation in attrition per wave is too large to be ascribed to the operation of a time-invariant first-order Markov process.

The chaos model (a first-order Markov chain in which the transition probabilities were allowed to vary per wave) accounts for all the variation in the data ($LR = 0$). However, it has no degrees of freedom and the model is not restrictive enough to be of any interest. All this model says is that a model in which a separate transition matrix is computed for every pair of successive waves fits the data perfectly – and the fact that randomness can also account for the data pattern is, of course, not a finding that greatly enhances our understanding of what is the case. Therefore, we must ask ourselves whether there is a more restricted (simpler) model that accounts acceptably well for the data.

The mover-stayer model with constant transition probabilities is such a simpler model. This model fits the data extremely well: LR is only 0.177 with two degrees of freedom; $p > 0.5$. Hence, in accordance with our *a priori* hypotheses, a two-chain model is able to reproduce the data very well. This would imply that the data set is composed of two groups, that differ concerning their response probabilities: the members of one group will never change their status (i.e., they will either always refuse to cooperate or always agree to cooperate). The other group consists of subjects whose attrition rate is identical and constant over time.

Finally, a two-chain model that did not restrict one of the transition matrices to be identical to the identity matrix was fitted (i.e., both chains consist of movers, but the groups may differ with respect to their transition matrices). Not surprisingly, this model hardly improves upon the mover-stayer model, and it has no degrees of freedom.

Which of these four models is best? The two simplest models that can

Table 3. Composition of the two chains over time

		Wave 1	Wave 2	Wave 3
Chain 1:	Participates	653	297	135
movers	Nonresponse	535	891	1053
($N = 1188$)				
Chain 2:	Participates	1122	1122	1122
stayers	Nonresponse	492	492	492
($N = 1614$)				
Total	Participates	1775	1419	1257
	Nonresponse	1027	1383	1545
	Total*	2802	2802	2802
Ratio members A to B		653/1775	297/1419	135/1122
Within participating subjects		(0.58)	(0.26)	(0.12)

* See Appendix: Total adds up to 2802 instead of 2800 for computational purposes.

account for the data are the mover-stayer model and the first-order Markov chain with time-varying response probabilities (the chaos model). Of these two models, the mover-stayer model is the simplest and theoretically the most attractive. Therefore, we prefer the mover-stayer model.[3]

The results for the mover-stayer model. The proportion of subjects in the first chain (movers) is 0.424 (1188 subjects), and the proportion in the second chain (stayers) is 0.576 (1614 subjects). Hence, 1614 subjects do not experience any change concerning their response status; the decreasing nonresponse rate is per definition due to changes that occur in the mover chain. About half (55.5%) of the movers initially participated in the first wave, while the remainder did not (the subjects in the 'mover' chain who did not participate in the first wave are *de facto* also stayers). For the second chain, 69.5 per cent participated in all waves, while the remainder (30.5 per cent, $N = 492$) did not participate in any wave.

The estimated transition matrix $\boldsymbol{O}_{ij}$ is

$$\begin{bmatrix} 0.455 & 0.545 \\ 0.001 & 0.999 \end{bmatrix}$$

for the movers and, obviously, the identity matrix for the stayers. The composition of the two subgroups of the sample at all waves is presented in Table 3.

Obviously, the composition of the stayer chain does not change over time. The composition of the mover chain, however, does. At every wave, about half of the subjects who participated in the previous wave drops out. At the third wave only 135 subjects still participated. It could be predicted that 0.455 of these subjects would participate in a follow-up wave, i.e., the model predicts that the total response for a fifth wave would be close to ((0.455×135) + 1122 equals) 1183 subjects, *ceteris paribus*.

The composition of the sample varies strongly per wave. At the first wave, 653 members of the mover chain participate, compared to 1122 members of the stayer chain (A to B-ratio is 0.58). This ratio is only 0.12 for the third wave. Note that the estimated ratio between movers and stayers in the population is (1188/1614 equals) 0.74, i.e., for every member of the mover chain there is 1.4 member of the stayer chain. For the sample as it was composed at the third wave, there are 8.3 stayers for each mover. Hence, one could say that at the third wave there are (8.3/1.4 equals) 5.9 times as many stayers as there should be in a representative sample. Obviously, there is cause for concern regarding the conclusions that are based on this sample.

This analysis, hence, shows that in the current example the attrition process can well be framed into the mover-stayer paradigm; two chains with different response probabilities. As outlined above. this may point to a lack of generalisability of the conclusions to the target population, at least, if one is willing to assume that the attrition rate is time-invariant. And we see no reason why it should not be: the fourteenth century philosopher and theologist William of Occam already stated that entities are not to be multiplied except as may be necessary (Occam's razor: simplicity above all). Why substitute the simple and theoretically plausible Markov model with a model without any restrictions on the transition probabilities? Of course, the chaos model could account for our findings; yet, we prefer the simpler explanation presented by the theoretically plausible two-chain Markov model.

5. Discussion

The current paper dealt with the issue of nonresponse in multiwave panel studies. In the first section it was shown that a decreasing nonresponse rate may well point at the operation of the dangerous process of selective nonresponse, i.e., response patterns can be used to diagnose heterogeneity in the sample. If subgroups are not the same concerning the probability that they will be (or remain) in the sample, this will result in a declining nonresponse rate.

In the following two sections it was discussed how discrete-time Markov models could be used to study nonresponse patterns. We have shown that it is possible to obtain some insight in the deglee to which the sample was stratified (i.e., determine the number of groups with different response probabilities, and these response probabilities themselves), and how to test for time variation in the rate of nonresponse. Such knowledge could be used to estimate the size of the sample in successive waves, and to assess the degree to which conclusions drawn on the basis of the sample can be generalised to the target population.

Our results can also be used to make several recommendations. First, we believe that an inspection of the pattern of nonresponse can yield important insights in the degree to which one's panel is still approximately representative for the target population, or that it could be severely biased. Even simply looking over this pattern, without any further statistical analyses, may reveal important problems with regard to the generalisability of results. As far as we are concerned, this simultaneously points at the duty of researchers to always report the nonresponse per wave of the study, and not just the total nonresponse: given the importance of this issue we feel this is not asking too much.

Second, our example has shown that selective nonresponse can rapidly lead to a severely biased sample. Therefore, this points at the absolute necessity to try to reduce nonresponse as much as possible: dropping out of the study should not be made easier than it already is, especially for the members of groups that have a low response probability. This is the only way to minimise the impact of selective nonresponse, though it should be noted that this impact can still be quite dramatic.

Further analysis

Having come to this point, some readers may wonder in what way the above discussion of nonresponse patterns relates to the common practice of linking biographic variables (such as level of education, gender and age) to nonresponse. We can think of two such relations. First, if gender, level of education, age, and so on, are not significantly correlated with the nonresponse and if the sample is about equal to the target population with respect to the distribution of these variables, it is still *not necessarily* the case that the sample is *representative* for the target population. Inspection of the nonresponse rate may reveal that the sample is biased, in spite of its superficial similarity to the target population. Hence, the procedure proposed above complements the existing techniques, and cannot replace these.

Second, let us assume that in a particular study, a decreasing nonresponse rate demonstrates that nonresponse was selective. *Post-hoc* analyses of the nonresponse reveal that men are overrepresented. Now the sample can be stratified according to gender, and analysed according to the procedures discussed above. If *within each stratum* the nonresponse is a *constant* (for example, for men a first-order Markov process with transition matrix $\boldsymbol{O}^m$ applies, while for women a different first-order Markov process with transition matrix $\boldsymbol{O}^w$ is the case), one has discovered the mechanism that generated the nonresponse. In subsequent research it is then possible to control the effects of nonresponse, for example, by weighting with gender.

If, however, the nonresponse within each stratum is still decreasing, the structure of the process responsible for the nonresponse has not yet completely been uncovered. Gender alone cannot account for the heterogeneity in the sample: indeed, the discrepancy between the cell frequencies as reproduced on the basis of the parameter estimates of the first-order Markov chain and the observed cell frequencies can be seen as an indication of the degree to which response-homogeneity is still questionable after controlling gender. If this difference is found unacceptable, one could proceed with stratification of the sample in yet other (smaller) sub-strata, until the discrepancy between a first-order Markov model and the data is acceptably small (see Taris, 1995, for an application).

Acknowledgements

The data used here were collected as part of a longitudinal study on the social integration of young adults in the Netherlands. This is a joint enterprise of the Department of Social Research Methods and the Department of Work and Organisational Psychology, both of the Free University of Amsterdam. The current research was supported by grant number PGS 56-381 of the Netherlands Organisation, for Scientific Research (NWO).

Appendix

*The data set used in the example (N = 2800)**

Wave 1	Wave 2	Wave 3	Frequency (absolute numbers)
Participated	Participated	Participated	1257
Participated	Participated	Nonresponse	162
Participated	Nonresponse	Nonresponse	356
Nonrespons	Nonresponse	Nonresponse	1025

* Note: In the analyses a wave (wave 1, 2, 3) × response (participated,

nonrespons) contingency table was analyzed. For computational purposes, 0.5 was added to the four cells in this eight-cell table that had a content of zero (i.e., all cells that involved a pattern in which 'nonrespons' was followed by 'participated', for example, 'nonrespons' at wave 1 and 'participated' at wave 2 and 3). Hence, in the analyses the total N amounts to 2802 instead of 2800.

Notes

1. The rate at which group B gets to dominate the sample is not only dependent on the difference in response rates, but also on the absolute size of the response rates. If $R_a = 0.8$ and $R_b = 0.7$, the A to B-ratio will increase much slower than in the case that $R_a = 0.2$ and $R_b = 0.1$.
2. This is not only because a first-order Markov process is far easier to discuss than a higher-order process, but also due to practical considerations: testing higher-order Markov models require many rounds of data collection, and longitudinal studies involving more than two or three waves are rare. To test a first-order Markov chain one needs at least three waves of data, to test a second-order Markov chain one needs at least four waves, and so on. An extension to higher-order Markov chains is possible, however: see Van de Pol and Langeheine, 1989).
3. One could contend that even the mover-stayer model counts many parameters compared to the number of data points, and, therefore, that such a model will always account for the data quite well. This is, however, not the case. For example, if the wave 1-wave 2 nonrespons would be 356 subjects and the wave 2-wave 3 nonresponse 162 (i.e., the nonresponse increases over time; compare the Appendix), the mover/stayer model would have a very poor fit (LR with two degrees of freedom is 111.92, $p < 0.001$).

References

Anderson, T.W. (1954). Probability models for analyzing time changes in attitudes, in P.F. Lazarsfeld (ed.), *Mathematical Thinking in the Social Sciences*, Glencoe, IL: The Free Press.

Bartholomew, D.J. (1981). *Mathematical Models in Social Science*, New York: Wiley.

Blumen, I., Kogan, M. & McCarthy, P.J. (1966). Probability models for mobility, in P.F. Lazarsfeld & N.W. Henry (eds.), *Readings in Mathematical Social Science*, Cambridge, MA: MIT Press, pp. 318–334.

Chung, K.L. (1967). *Markov Chains with Stationary Transition Matrices*, Berlin: Springer.

Coleman, J.S. (1968). The mathematical study of change, in H.M. Blalock & A.B. Blalock (eds.), *Methodology in Social Research*,New York: McGraw-Hill, pp. 428–478.

Dijkstra, W.D. (ed.) (1993). *Het proces van sociale integratie van jong-volwassenen* (The process of social integration of young adults), Amsterdam: VU-Uitgeverij.

Freedman, D. (1970). *Markov Chains*, San Francisco: Holden-Day.

Goyder, J. (1987). *The Silent Minority*, Cambridge: Blackwell University Press.

Hox, J.J. & De Leeuw, E.D. (1993). Nonrespons bij schriftelijke vragenlijsten, telefonische interviews en face-to-face interviews: een vergelijkende meta-analyse (Nonresponse in mail surveys, telephone interviews and face-to-face interviews: A comparative meta-analysis), *Kwantitatieve Methoden*, 14, 59–78.

Langeheine, R. & Van de Pol, F.J.R. (1989). A unifying framework for Markov modeling in discrete space and discrete time, *Sociological Methods & Research*, 18.

Markus, G.B. (1979). *Analyzing Panel Data*, Beverly Hills: Sage.

Poulsen, C.S. (1982). *Latent Structure Analysis with Choice Modeling Applications*, Philadelphia: Wharton School (Unpublished PhD-dissertation).

Taris, T.W. (1994). *Representativiteit en nonresponse in panel studies: een korte notitie* (Representativity and nonresponse in panel studies: A short note). *Kwantitatieve Methoden*, 16, 41–44.

Taris, T.W. (1995). *On Selectivity of Nonresponse in Discrete-Time Multi-Wave Panel Studies.* Manuscript submitted for publication.

Taris, T.W., Van der Vaart, W. & Dijkstra, W.D. (1993). Nonrespons en de representativiteit van de follow-up metingen (Nonresponse and representativity of the follow-up waves), in W.D. Dijkstra (ed.), *Het proces van sociale integratie van jong-volwassenen*, Amsterdam: VU-Uitgeverij, pp. 65–76.

Van de Pol, F.J.R. (1989). *Issues of Design and Analysis of Panels*, Amsterdam: Sociometric Research Foundation.

Van de Pol, F.J.R., De Jong, W. & Langeheine, R. (1990). *PANMARK 2.2 User Manual: Panel Analysis using Markov Chains*, Voorburg: Netherlands Central Bureau of Statistics CBS.

Van de Pol, F.J.R. & Langeheine, R. (1989). Mixed Markov latent class models, from description toward explanation, in C.C. Clog (ed.), *Sociological Methodology 1990.*

Wiggins, L.M. (1973). *Panel Analysis, Latent Probability Models for Attitude and Behavior Processes*, Amsterdam: Elsevier.